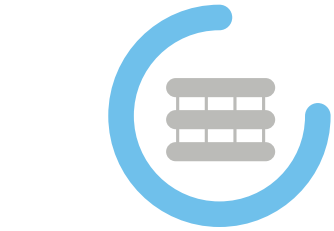


Meeting of the Technical Steering Committee (TSC) Board

Wednesday, May 06th, 2020
11:00am ET

Meeting Logistics

- https://zoom.us/j/556149142

- United States : +1 (646) 558-8656
  - Meeting ID: 556 149 142

Antitrust Policy Notice

- Linux Foundation meetings involve participation by industry competitors, and it is the intention of the Linux Foundation to conduct all of its activities in accordance with applicable antitrust and competition laws. It is therefore extremely important that attendees adhere to meeting agendas, and be aware of, and not participate in, any activities that are prohibited under applicable US state, federal or foreign antitrust and competition laws.

- Examples of types of actions that are prohibited at Linux Foundation meetings and in connection with Linux Foundation activities are described in the Linux Foundation Antitrust Policy available at http://www.linuxfoundation.org/antitrust-policy. If you have questions about these matters, please contact your company counsel, or if you are a member of the Linux Foundation, feel free to contact Andrew Updegrove of the firm of Gesmer Updegrove LLP, which provides legal counsel to the Linux Foundation.

THE LINUX FOUNDATION

Agenda/Updates

- Announcements:
  - ISC in-person event cancelled – will be a digital event instead
    - 4/24: Bofs have been cancelled entirely
    - "*We will ask for a confirmation to present your BoF at ISC 2021 in a few months*"
- Upcoming deadlines:
  - SC'20
    - Tutorials?: Due ~~April 16, 2020~~, Extended to May 08 – David B. taking over from Adrian for submission)
    - BoFs: Due July 31, 2020

- TSC Nominations
- PEARC'20 tutorial/cloud working group updates (csim)
- SLURM config updates
- Conflicts impacting multiple compiler vendor installs
- UCX/libfabric
- 2.0 build status

TSC Selections

- We are coming up on the end of the current 2019-2020 TSC service cycle (term ends June 30)
- Plan to post call for nominations as usual to the openhpc-users list this week
- Reminder on process and timeline
  - expect to follow same schedule as last year to have 2020-2021 TSC selections completed by July 1
  - Submission deadline: June 12th, 2020 (Friday)

TSC Selections

- Reminder on the overall process:
  - **anyone** can submit nominations (including **self-nominations** by the June 12th deadline)
  - submit nomination and CV to following email: tsc-nominations@OpenHPC.groups.io
  - When nominating, please highlight any potential conflicts of interest and indicate the role you are primarily interested in:
    - maintainer
    - end-user site representative
    - upstream component development representative
    - testing coordinator
  - After the nomination deadline, current TSC members will receive a request to vote on next year's nominees via a Google form
    - current TSC members have until June 30 to cast their votes
- New TSC membership for 2020-2021 term announced July 1st
- With new TSC membership in place, next step is to take nominations for TSC Project Lead (from pool of TSC members)
  - nominations will go to Neal
  - TSC members will then receive another Google form to cast votes
  - Project Lead selection process to be completed by July 31, 2020
- Any questions?

OHPC Cloud Working Group Updates (csim)

- Virtual PEARC technical problems presents unique challenges
- Shifting focus from primary product being a presentation
- Instead, product will be self-directed educational content
  - We will run through the lesson during the conference
  - Aim to provide
    - some AWS hours for others to do after
    - 'office hours' via slack or other means
    - access to AWS based system as user to explore
  - Simmons and students get back to it next week (after finals)
  - Downing has checked in stripped down scripts to start from

SLURM config updates

- 2.0 slurm build has been updated to the latest 20.x series
  - we are on 18.x series in ohpc 1.3.9
- There are some new features and changes to defaults to be aware of:
  1. **exclusive** scheduling not enabled out of the box
     - we have relied on no node sharing and assume that is preferred for HPC systems out of the box
     - user's could always opt-in to share resources on a node if they wanted
     - latest build was allowing multiple MPI jobs on a node during CI
     - consequently our slurm recipe has been updated to request exclusive usage:

```
PartitionName=normal Nodes=c[1-4] Default=YES MaxTime=24:00:00 State=UP Oversubscribe=EXCLUSIVE
```

  2. "**Configless**" option
     - allows compute nodes to pull config from slurmctld directly
     - means you don't have to keep a slurm.conf file in sync on all computes/logins
     - https://slurm.schedmd.com/configless_slurm.html
     - Have updated slurm recipes to take advantage of this

```
SlurmctldParameters=enable_configless
```

← in slurm.conf on master host

```
# cat /etc/sysconfig/slurmd
SLURMD_OPTIONS=--conf-server 192.168.1.5
```

← setting on compute host to point to master controller

SLURM config updates

- How do you know if configless setup is working?
  - On a configless host, copy of slurm.conf is still pulled down
    - resides in `/var/run/slurm/conf/slurm.conf`
  - You can make changes on master controller and push the changes as follows:

```
[root@sms005 ~]# ssh c1 "tail -2 /var/run/slurm/conf/slurm.conf"
HealthCheckProgram=/usr/sbin/nhc
HealthCheckInterval=300
```

← examine state of config on compute host

```
[root@sms005 ~]# echo "# a superfly comment" >> /etc/slurm/slurm.conf
[root@sms005 ~]# scontrol reconfig
```

← make noop change on master controller and push

```
[root@sms005 ~]# ssh c1 "tail -2 /var/run/slurm/conf/slurm.conf"
HealthCheckInterval=300
# a superfly comment
```

← reexamine state of config on compute host

Package conflicts impacting multiple compiler vendor installs

- Our CI recipes for including 3rd party builds for the intel compiler occur after install of gcc variants
- Part of the reason we have been seeing poor test coverage with (gcc + intel) is due to file conflict preventing package installs

```
file /usr/lib/.build-id/39/be892dc0585c6a5a81c74ec63711df8e9b11e9 from install of pdtoolkit-intel-ohpc-3.25.1-3.1.ohpc.2.0.x86_64 conflicts with file from package pdtoolkit-gnu9-ohpc-3.25.1-2.1.ohpc.2.0.x86_64

file /usr/lib/.build-id/b8/f0f92f577673f2d23893cc76cd7cbb94f6592d from install of pdtoolkit-intel-ohpc-3.25.1-3.1.ohpc.2.0.x86_64 conflicts with file from package pdtoolkit-gnu9-ohpc-3.25.1-2.1.ohpc.2.0.x86_64
```

- The default rpmbuild setup for rhel8 generates these .build-id files – intended for use with debuginfo packages
- Have updated our global OHPC_macros file to disable this option which should resolve a number of package installs…

```
# Disable generation of .build-id links
%global _build_id_links none
```

libfabric – update regarding distro version(s)

- Last time, we agreed to include libfabric (for relevant MPI builds) but rely on distro version if possible
- Have completed an MPICH build with distro libfabric
- Works (mostly) fine on ethernet but have been unable to get running on IB network
  - CI outliers are geopm, miniFE, FFTW, and PETSc
- Digging further it seems there is an issue with current rhel distro version which may force us to rethink our position
  - Consider CI system with MLNX HCA and active IB connection

```
karl@sms001 ~ $ ibv_devinfo
hca_id:         mlx4_0
        transport:                      InfiniBand (0)
        fw_ver:                         2.36.5000
        node_guid:                      001e:6703:00f2:54d2
        sys_image_guid:                 001e:6703:00f2:54d5
        vendor_id:                      0x02c9
        vendor_part_id:                 4099
        hw_ver:                         0x1
        board_id:                       INCX-3I355920151
        phys_port_cnt:                  1
                port:   1
                        state:          PORT_ACTIVE (4)
                        max_mtu:        4096 (5)
                        active_mtu:     4096 (5)
                        sm_lid:         1
                        port_lid:       5
                        port_lmc:       0x00
                        link_layer:     InfiniBand
```

libfabric

- Had initial issue with building against distro provided libfabric
- MPICH was still building against it's own internal copy
- Have prepared an ohpc-build of libfabric just in case along with mpich build using our libfabric
  - Example of transport providers available on our (x86) CI system

```
karl@sms001 ~ $ which fi_info
/opt/ohpc/pub/mpi/libfabric/1.10.0/bin/fi_info

karl@sms001 ~ $ fi_info  | grep provider
provider: verbs
provider: tcp;ofi_rxm
provider: tcp;ofi_rxm
provider: tcp;ofi_rxm
provider: tcp;ofi_rxm
provider: verbs;ofi_rxd
provider: UDP;ofi_rxd
provider: UDP;ofi_rxd
provider: UDP;ofi_rxd
provider: UDP;ofi_rxd
provider: shm
provider: UDP
provider: UDP
provider: UDP
provider: UDP
provider: tcp
provider: tcp
provider: tcp
provider: tcp
provider: sockets
provider: sockets
provider: sockets
provider: sockets
```

libfabric + mpich: example runs

[default run]

```
# OSU MPI Bandwidth Test v5.6.2
# Size      Bandwidth (MB/s)
1                         0.26
2                         0.51
4                         0.99
8                         1.76
16                        4.05
32                        4.76
64                        8.78
128                      21.32
256                      34.83
512                      54.61
1024                     77.38
2048                     85.90
4096                    105.08
8192                    109.41
16384                   113.46
32768                   114.91
65536                   116.23
131072                  116.64
262144                  117.27
524288                  117.46
1048576                 117.58
```

[explicitly choose IB provider]

```
karl@c1 ~ $ export FI_PROVIDER=verbs


# OSU MPI Bandwidth Test v5.6.2
# Size      Bandwidth (MB/s)
1                         1.26
2                         2.57
4                         5.20
8                        10.58
16                       21.09
32                       42.15
64                       80.59
128                     161.45
256                     274.65
512                     508.20
1024                    916.57
2048                   1682.75
4096                   2276.75
8192                   2543.37
16384                  1775.88
32768                  2505.65
65536                  2948.42
131072                 3205.82
262144                 3272.77
524288                 3230.84
1048576                3123.51
```

UCX

- ohpc build of UCX is now in 2.0 factory
- also have an MPICH variant that uses our UCX build (requires use of ch4)
- ucx-ohpc package:
  - First example of transport providers available on our (x86) CI system with install of ucx-ohpc
  - Second example after installing additional ucx provider (for IB)

```
karl@sms001 ~ $ module list ucx

Currently Loaded Modules Matching: ucx
  1) ucx/1.8.0

karl@sms001 ~ $ ucx_info  -d | grep Transport
#   Transport: posix
#   Transport: sysv
#   Transport: self
#   Transport: tcp
```

```
[root@sms001 SPECS]# yum install ucx-ib-ohpc


karl@sms001 ~ $ ucx_info  -d | grep Transport
#   Transport: posix
#   Transport: sysv
#   Transport: self
#   Transport: tcp
#   Transport: rc_verbs
#   Transport: ud_verbs
```

UCX + mpich: example runs

[default run prior to install of ucx-ib-ohpc]

```
# OSU MPI Bandwidth Test v5.6.2
# Size      Bandwidth (MB/s)
1                         0.41
2                         0.77
4                         1.47
8                         2.72
16                        4.85
32                       10.00
64                       18.09
128                      25.55
256                      52.16
512                      75.43
1024                     90.60
2048                    101.64
4096                    107.20
8192                    114.28
16384                   115.01
32768                   116.32
65536                   116.84
131072                  117.25
262144                  117.46
524288                  117.54
1048576                 117.60
2097152                 117.65
4194304                 117.67
```

[default run after install of ucx-ib-ohpc]

```
# OSU MPI Bandwidth Test v5.6.2
# Size      Bandwidth (MB/s)
1                         4.99
2                         9.73
4                        19.98
8                        38.89
16                       80.14
32                      159.81
64                      272.48
128                     443.47
256                     883.28
512                    1722.72
1024                   3027.96
2048                   4277.59
4096                   5139.35
8192                   5600.89
16384                  5766.42
32768                  5850.02
65536                  5906.69
131072                 5923.59
262144                 6364.84
524288                 6372.58
1048576                6380.71
2097152                6381.43
4194304                6381.75
```

Comparison runs for IB

[MPICH/libfabric]

```
# OSU MPI Bandwidth Test v5.6.2
# Size      Bandwidth (MB/s)
1                       1.26
2                       2.57
4                       5.20
8                      10.58
16                     21.09
32                     42.15
64                     80.59
128                   161.45
256                   274.65
512                   508.20
1024                  916.57
2048                 1682.75
4096                 2276.75
8192                 2543.37
16384                1775.88
32768                2505.65
65536                2948.42
131072               3205.82
262144               3272.77
524288               3230.84
1048576              3123.51
2097152              3121.35
4194304              3089.71
```

[MPICH/UCX]

```
# OSU MPI Bandwidth Test v5.6.2
# Size      Bandwidth (MB/s)
1                       4.99
2                       9.73
4                      19.98
8                      38.89
16                     80.14
32                    159.81
64                    272.48
128                   443.47
256                   883.28
512                  1722.72
1024                 3027.96
2048                 4277.59
4096                 5139.35
8192                 5600.89
16384                5766.42
32768                5850.02
65536                5906.69
131072               5923.59
262144               6364.84
524288               6372.58
1048576              6380.71
2097152              6381.43
4194304              6381.75
```

[MVAPICH2]

```
# OSU MPI Bandwidth Test v5.6.2
# Size      Bandwidth (MB/s)
1                       5.01
2                       9.88
4                      20.20
8                      39.37
16                     78.13
32                    151.93
64                    326.92
128                   655.60
256                  1342.96
512                  2287.98
1024                 3418.03
2048                 4387.52
4096                 5064.07
8192                 5643.30
16384                5960.62
32768                6169.67
65536                6256.33
131072               6296.92
262144               6340.58
524288               6356.74
1048576              6371.37
2097152              6346.07
4194304              6381.23
```

mpich: ucx/libfabric

- I do not believe we can have a monolithic build that is capable of using both ucx an libfabric
  - have reached out to mpich contact at Argonne to hopefully confirm/deny
- Assuming that is the case, do we want to provide variants for both?
  - would one be assumed default choice? this is sort of what we do for pmiX
  - or, would you call out variants directly and force user to choose?
    - mpich-gnu9-ucx-ohpc
    - mpich-gnu9-ofi-ohpc
- Do folks have insight into long-term investment/viability of both?
  - full test coverage of both obviously requires significant time

2.0 Build Status

- Have enabled a good number of packages in 2.0 Factory
  - introduction of arm1 compiler variant means 10 compiler/MPI permutations for MPI-based builds

Administrative Tools

Package	Built?	Notes
conman	✓	
docs	✓	
examples	✓	
~~ganglia~~		
genders	✓	
lmod-defaults	✓	added (3/10/20)
losf	✓	
mrsh	✓	fixed (3/8/20)
~~nagios~~		
~~nagios-plugins~~		
~~ndoutils~~		
nhc	✓	added (3/10/20)
~~nrpe~~		
pdsh	✓	
prun	✓	
test-suite	✓	

Compiler Families

Package	Built?	Notes
gcc9	✓	
intel-compatibility	✓	
arm-compatibility	✓	
llvm		

Development Tools

Package	Built?	Notes
easybuild	✓	
autoconf	✓	
automake	✓	
cmake	✓	
hwloc	✓	
libtool	✓	
python-mpi4py	※	3/24/20 – arm1 failures
python-numpy	※	3/24/20 – arm1 failures
python-scipy	※	3/19/20 – arm1 failures
spack	✓	[still targets root usage though]
valgrind	✓	

MPI Families

Package	Built?	Notes
impi-compatibility	✓	no longer requires install of intel compat. package
mpich	✓	
mvapich2	✓	
openmpi4	✓	

Serial Libs

Package	Built?	Notes
R	✓	
GSL	✓	
metis	✓	
openblas	✓	
plasma	✓	arm1 enabled (4/21/20)
superlu	✓	arm1 fixed (3/9/20)

※ = partial builds success
✓ = all builds complete
X = all builds fail

2.0 Build Status (cont.)

Parallel Libraries

Package	Built?	Notes
boost	※	intel error
fftw	✓	build fixed
hypre	※	intel/arm1 blas issue
mfem	※	needs deps
mumps	※	arm1/impi issue
opencoarrays	✓	cmake issue fixed (5/3)
petsc	※	intel/arm1 issue
scotch	✓	arm1 fixed (3/9/20)
scalapack	✓	arm1 fixed (3/9/20)
slepc	※	waiting on petsc deps
superlu_dist	※	2 Leap 15.1 failures
trilinos	※	Intel/ARM failures (4/8/20)

Resource Management

Package	Built?	Notes
PBS Pro	※	need libical-devel on aarch/centos
pmix	✓	
slurm	✓	

Performance Tools

Package	Built?	Notes
dimemas	※	need some boost builds, suse failures
extrae	※	arm1 failures
geopm	※	intel failures (4/18/20)
imb	※	updated to v2019.6, all built except for arm1 failures (4/16/20)
likwid	✓	(4/7/20)
msr-safe	✓	(4/20/20)
~~mpiP~~		
omb	✓	3/4/20
papi	✓	
paraver	✓	added 3/5/20
pdtoolkit	※	intel/leap failure (3/5/20)
scalasca	※	need scorep deps (3/5/20)
scorep	※	variety of issues (3/5/20)
tau	※	need pdtoolkit on intel/leap and arm1 (3/24/20)

Runtimes

Package	Built?	Notes
charliecloud	✓	3/23/20
~~ocr~~		
singularity	✓	updated to use newer go version (4/6/20)

※ = partial builds success
✓ = all builds complete
X = all builds fail

2.0 Build Status (cont.)

I/O Libraries

Package	Built?	Notes
adios	※	intel failures (and deps)
hdf5	※	Leap/arm1 failures
netcdf-cxx	※	Leap/arm1 failures
netcdf-fortran	※	Leap/arm1 failures
phdf5	※	arm1 failures
pnetcdf	✓	
sionlib	※	arm1 failures

Provisioning

Package	Built?	Notes
warewulf	✓	starting to run in CI (3/10/20)

※ = partial builds success
✓ = all builds complete
X = all builds fail

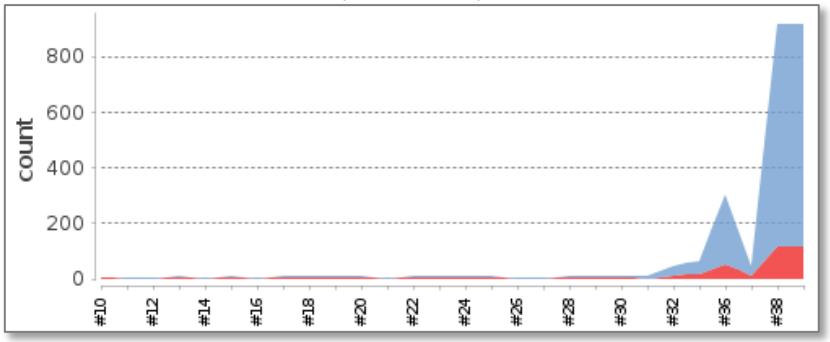
2.0 Build Status (cont)

- Current package counts:
  - ~~800 RPMs as of 3/11/2020~~
  - ~~905 RPMs as of 3/25/2020~~
  - ~~982 RPMs as of 4/8/2020~~
  - ~~1003 RPMs as of 4/22/2020~~
  - 1021 RPMs as of 5/6/2020

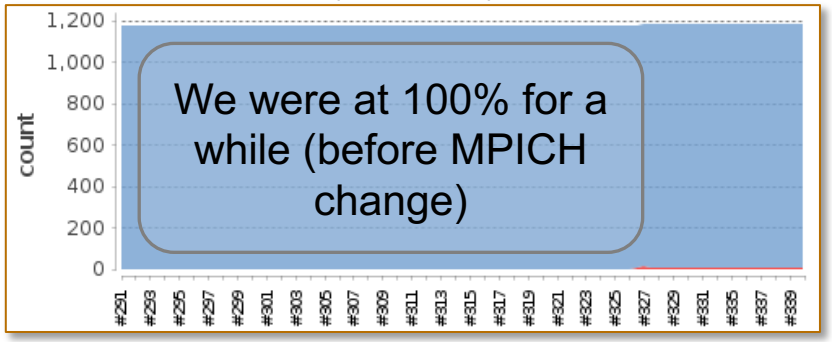
Base OS	aarch64	x86_64	noarch
CentOS 8	169	320	29
Leap 15	167	307	29

2.0 (cont.) – CI updates

centos8.1/x86/WW/slurm (3/10/2020)



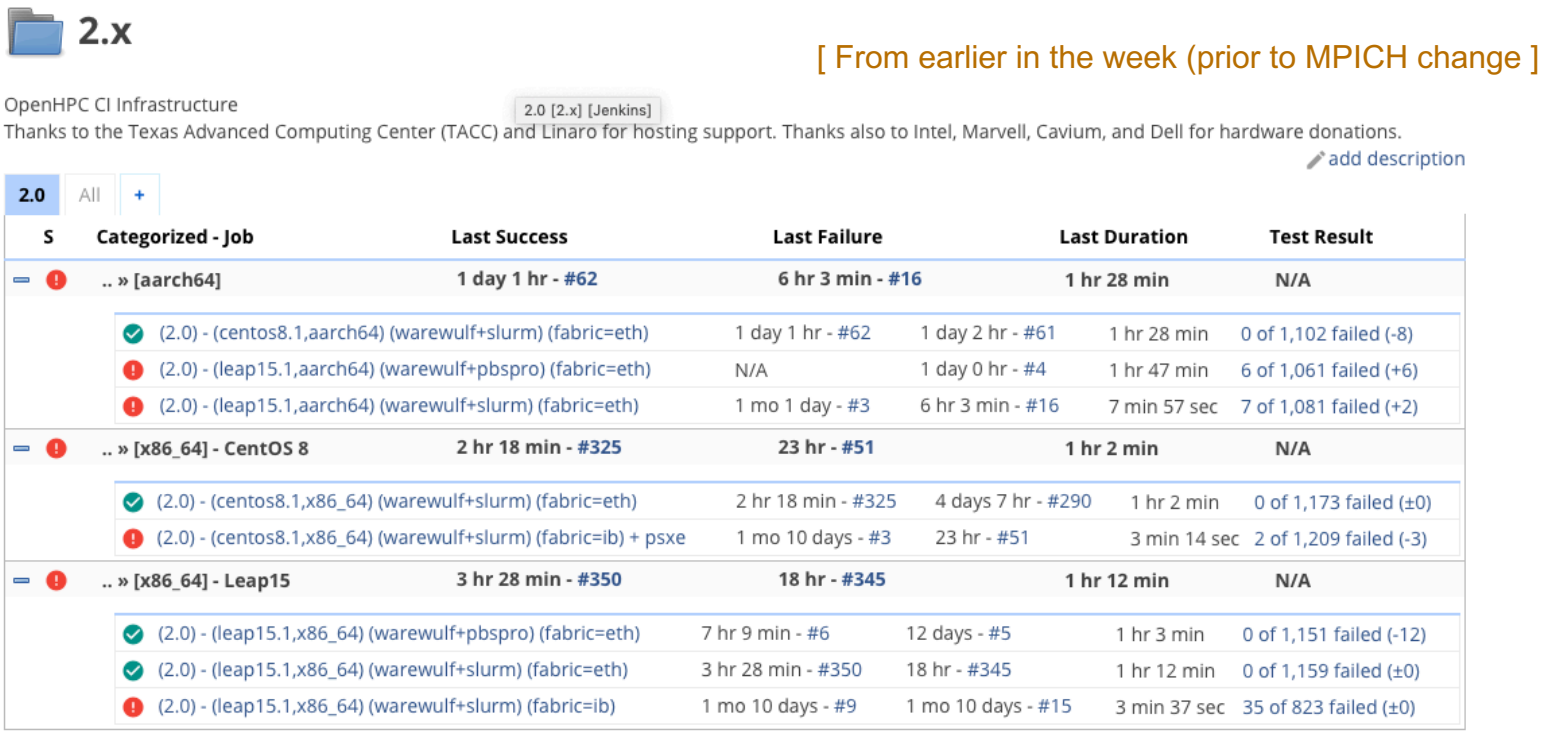
centos8.1/x86/WW/slurm (05/06/2020)



- As of 3/10:
  - ~~610 user-level tests passing~~
- As of 3/25:
  - ~~910 user-level tests passing~~
- As of (4/08):
  - ~~982 user-level tests passing~~
- As of (4/22):
  - ~~1,132 user-level tests passing~~

- As of this morning (4/22):
  - 1,140 user-level tests passing

2.0 (cont.) – CI updates



2.x

[From earlier in the week (prior to MPICH change]

OpenHPC CI Infrastructure

2.0 [2.x] [Jenkins]

Thanks to the Texas Advanced Computing Center (TACC) and Linaro for hosting support. Thanks also to Intel, Marvell, Cavium, and Dell for hardware donations.

add description

2.0 | All | +

S	Categorized - Job	Last Success	Last Failure	Last Duration	Test Result
	.. » [aarch64]	1 day 1 hr - #62	6 hr 3 min - #16	1 hr 28 min	N/A
	(2.0) - (centos8.1,aarch64) (warewulf+slurm) (fabric=eth)	1 day 1 hr - #62	1 day 2 hr - #61	1 hr 28 min	0 of 1,102 failed (-8)
	(2.0) - (leap15.1,aarch64) (warewulf+pbspro) (fabric=eth)	N/A	1 day 0 hr - #4	1 hr 47 min	6 of 1,061 failed (+6)
	(2.0) - (leap15.1,aarch64) (warewulf+slurm) (fabric=eth)	1 mo 1 day - #3	6 hr 3 min - #16	7 min 57 sec	7 of 1,081 failed (+2)
	.. » [x86_64] - CentOS 8	2 hr 18 min - #325	23 hr - #51	1 hr 2 min	N/A
	(2.0) - (centos8.1,x86_64) (warewulf+slurm) (fabric=eth)	2 hr 18 min - #325	4 days 7 hr - #290	1 hr 2 min	0 of 1,173 failed (±0)
	(2.0) - (centos8.1,x86_64) (warewulf+slurm) (fabric=ib) + psxe	1 mo 10 days - #3	23 hr - #51	3 min 14 sec	2 of 1,209 failed (-3)
	.. » [x86_64] - Leap15	3 hr 28 min - #350	18 hr - #345	1 hr 12 min	N/A
	(2.0) - (leap15.1,x86_64) (warewulf+pbspro) (fabric=eth)	7 hr 9 min - #6	12 days - #5	1 hr 3 min	0 of 1,151 failed (-12)
	(2.0) - (leap15.1,x86_64) (warewulf+slurm) (fabric=eth)	3 hr 28 min - #350	18 hr - #345	1 hr 12 min	0 of 1,159 failed (±0)
	(2.0) - (leap15.1,x86_64) (warewulf+slurm) (fabric=ib)	1 mo 10 days - #9	1 mo 10 days - #15	3 min 37 sec	35 of 823 failed (±0)

- We are unfortunately now down 2 of our Jenkins aarch64 nodes in the UK
  - one host throwing h/w error during bios post (hardware: needs riser reset)
  - <new host lost BMC access>

2.0 (cont.) – CI updates

[Regressions after MPICH change to use ch4:ucx transport]